

M. Ibrahim

1638

RAW SEQUENCE LISTING  
 PATENT APPLICATION: US/09/147,955

DATE: 05/24/2000  
 TIME: 10:58:34

Input Set : A:\1560-350.app  
 Output Set: N:\CRF3\05242000\I147955.raw

ENTERED

4 <110> APPLICANT: MIZUTANI, Masako  
 5 TANAKA, Yoshikazu  
 6 KUSUMI, Takaaki  
 7 SAITO, Kazuki  
 8 YAMAZAKI, Mami  
 9 ZHIZHONG, Gong  
 11 <120> TITLE OF INVENTION: GENES ENCODING PROTEINS HAVING TRANSGLYCOSYLATION  
 12 ACTIVITY  
 14 <130> FILE REFERENCE: 001560-350  
 16 <140> CURRENT APPLICATION NUMBER: 09/147,955  
 17 <141> CURRENT FILING DATE: 1999-03-24  
 19 <150> PRIOR APPLICATION NUMBER: PCT/JP98/03199  
 20 <151> PRIOR FILING DATE: 1998-07-16  
 22 <150> PRIOR APPLICATION NUMBER: JP 9-200571  
 23 <151> PRIOR FILING DATE: 1997-07-25  
 25 <160> NUMBER OF SEQ ID NOS: 12  
 27 <170> SOFTWARE: PatentIn Ver. 2.0  
 29 <210> SEQ ID NO: 1  
 30 <211> LENGTH: 1507  
 31 <212> TYPE: DNA  
 32 <213> ORGANISM: Perilla frutescens  
 34 <220> FEATURE:  
 35 <221> NAME/KEY: CDS  
 36 <222> LOCATION: (17)..(1396)  
 38 <400> SEQUENCE: 1  
 39 gaaaatttcc acaaaa atg gtc cgc cgc cgc gtc ctg cta gca acg ttt cct 52  
 40 Met Val Arg Arg Arg Val Leu Leu Ala Thr Phe Pro  
 41 1 5 10  
 43 gcg caa ggc cac ata aat ccc gcc ctc caa ttc gcc aag aga ctc cta 100  
 44 Ala Gln Gly His Ile Asn Pro Ala Leu Gln Phe Ala Lys Arg Leu Leu  
 45 15 20 25  
 47 aaa gcc ggc act gac gtc aca ttt ttc acg agc gtt tat gca tgg cgc 148  
 48 Lys Ala Gly Thr Asp Val Thr Phe Phe Thr Ser Val Tyr Ala Trp Arg  
 49 30 35 40  
 51 cgc atg gcc aac aca gcc tcc gcc gct gcc gga aac cca ccg ggc ctc 196  
 52 Arg Met Ala Asn Thr Ala Ser Ala Ala Ala Gly Asn Pro Pro Gly Leu  
 53 45 50 55 60  
 55 gac ttc gtg gcg ttc tcc gac ggc tac gac gac ggg ctg aag ccc tgc 244  
 56 Asp Phe Val Ala Phe Ser Asp Gly Tyr Asp Asp Gly Leu Lys Pro Cys  
 57 65 70 75  
 59 ggc gac ggg aag cgc tac atg tcc gag atg aaa gcc cgc ggc tcc gag 292  
 60 Gly Asp Gly Lys Arg Tyr Met Ser Glu Met Lys Ala Arg Gly Ser Glu  
 61 80 85 90  
 63 gcc tta aga aac ctc ctt ctc aac aac cac gac gtc acg ttc gtc gtc 340  
 64 Ala Leu Arg Asn Leu Leu Leu Asn Asn His Asp Val Thr Phe Val Val  
 65 95 100 105  
 67 tac tcc cac ctc ttt gca tgg gcg gcg gag gtg gcg cgt gag tcc cag 388

RECEIVED

JUN -9 2000

TC 1600 MAIL ROOM

RAW SEQUENCE LISTING  
 PATENT APPLICATION: US/09/147,955  
 DATE: 05/24/2000  
 TIME: 10:58:34

Input Set : A:\1560-350.app  
 Output Set: N:\CRF3\05242000\I147955.raw

```

68 Tyr Ser His Leu Phe Ala Trp Ala Ala Glu Val Ala Arg Glu Ser Gln
69      110      115      120
71 gtc ccg agc gcc ctt ctc tgg gtc gag ccc gcc acc gtg ctg tgc ata 436
72 Val Pro Ser Ala Leu Leu Trp Val Glu Pro Ala Thr Val Leu Cys Ile
73 125      130      135      140
75 tat tac ttc tac ttc aac ggc tac gca gac gag atc gac gcc ggt tcc 484
76 Tyr Tyr Phe Tyr Phe Asn Gly Tyr Ala Asp Glu Ile Asp Ala Gly Ser
77      145      150      155
79 gac gaa att cag ctc cct cgg ctt cca ccc ctg gag cag cgc agt ctt 532
80 Asp Glu Ile Gln Leu Pro Arg Leu Pro Pro Leu Glu Gln Arg Ser Leu
81      160      165      170
83 ccg acc ttt ctg ctg ccg gag aca ccg gag aga ttc ccg ttg atg atg 580
84 Pro Thr Phe Leu Leu Pro Glu Thr Pro Glu Arg Phe Arg Leu Met Met
85      175      180      185
87 aag gag aag ctg gaa act tta gac ggt gaa gag aag gcg aaa gtg ttg 628
88 Lys Glu Lys Leu Glu Thr Leu Asp Gly Glu Glu Lys Ala Lys Val Leu
89      190      195      200
91 gtg aac acg ttt gat gcg ttg gag ccc gat gca ctc acg gct att gat 676
92 Val Asn Thr Phe Asp Ala Leu Glu Pro Asp Ala Leu Thr Ala Ile Asp
93 205      210      215      220
95 agg tat gag ttg atc ggg atc ggg ccg ttg att ccc tcc gcc ttc ttg 724
96 Arg Tyr Glu Leu Ile Gly Ile Gly Pro Leu Ile Pro Ser Ala Phe Leu
97      225      230      235
99 gac ggc gga gat ccc tcc gaa acg tct tac ggc ggc gat ctt ttc gaa 772
100 Asp Gly Gly Asp Pro Ser Glu Thr Ser Tyr Gly Gly Asp Leu Phe Glu
101      240      245      250
103 aaa tcg gag gag aat aac tgc gtg gag tgg ttg gac acg aag ccg aaa 820
104 Lys Ser Glu Glu Asn Asn Cys Val Glu Trp Leu Asp Thr Lys Pro Lys
105      255      260      265
107 tct tcg gtg gtg tat gtg tgc ttt ggg agc gtt ttg agg ttt cca aag 868
108 Ser Ser Val Val Tyr Val Ser Phe Gly Ser Val Leu Arg Phe Pro Lys
109      270      275      280
111 gca caa atg gaa gag att ggg aaa ggg cta tta gcc tgc gga agg ccg 916
112 Ala Gln Met Glu Glu Ile Gly Lys Gly Leu Leu Ala Cys Gly Arg Pro
113 285      290      295      300
115 ttt tta tgg atg ata cga gaa cag aag aat gac gac ggc gaa gaa gaa 964
116 Phe Leu Trp Met Ile Arg Glu Gln Lys Asn Asp Asp Gly Glu Glu Glu
117      305      310      315
119 gaa gaa gag ttg agt tgc att ggg gaa ttg aaa aaa atg ggg aaa ata 1012
120 Glu Glu Glu Leu Ser Cys Ile Gly Glu Leu Lys Lys Met Gly Lys Ile
121      320      325      330
123 gtt tcg tgg tgc tgc cag ttg gag gtt ctg gcg cac cct gcg ttg gga 1060
124 Val Ser Trp Cys Ser Gln Leu Glu Val Leu Ala His Pro Ala Leu Gly
125      335      340      345
127 tgt ttc gtg acg cat tgt ggg tgg aac tcg gct gtg gag agc ttg agt 1108
128 Cys Phe Val Thr His Cys Gly Trp Asn Ser Ala Val Glu Ser Leu Ser
129      350      355      360
131 tgc ggg gtt ccg gtg gtg gcg gtg ccg cag tgg ttt gat cag acg acg 1156
132 Cys Gly Val Pro Val Val Ala Val Pro Gln Trp Phe Asp Gln Thr Thr

```

## RAW SEQUENCE LISTING

DATE: 05/24/2000

PATENT APPLICATION: US/09/147,955

TIME: 10:58:35

Input Set : A:\1560-350.app

Output Set: N:\CRF3\05242000\I147955.raw

```

133 365          370          375          380
135 aat gcg aag ctg att gag gat gcg tgg ggg aca ggg gtg aga gtg aga 1204
136 Asn Ala Lys Leu Ile Glu Asp Ala Trp Gly Thr Gly Val Arg Val Arg
137          385          390          395
139 atg aat gaa ggg ggt ggg gtt gat gga tct gag ata gag agg tgt gtg 1252
140 Met Asn Glu Gly Gly Gly Val Asp Gly Ser Glu Ile Glu Arg Cys Val
141          400          405          410
143 gag atg gtg atg gat ggg ggt gag aag agc aaa cta gtg aga gaa aat 1300
144 Glu Met Val Met Asp Gly Gly Glu Lys Ser Lys Leu Val Arg Glu Asn
145          415          420          425
147 gcc ata aaa tgg aag act ttg gcc aga gaa gcc atg gga gag gat gga 1348
148 Ala Ile Lys Trp Lys Thr Leu Ala Arg Glu Ala Met Gly Glu Asp Gly
149          430          435          440
151 tct tca ctc aag aat ctc aac gcc ttt ctt cat caa gtt gca cgt gct 1396
152 Ser Ser Leu Lys Asn Leu Asn Ala Phe Leu His Gln Val Ala Arg Ala
153 445          450          455          460
155 taatacacaa aatggctttc cacttttaat ctactcaaac accggttcaa ataaatatcc 1456
157 ccttcacatt ctttctattt cactatcaca ttataattt tagtaacaaa a 1507
160 <210> SEQ ID NO: 2
161 <211> LENGTH: 460
162 <212> TYPE: PRT
163 <213> ORGANISM: Perilla frutescens
165 <400> SEQUENCE: 2
166 Met Val Arg Arg Arg Val Leu Leu Ala Thr Phe Pro Ala Gln Gly His
167 1 5 10 15
169 Ile Asn Pro Ala Leu Gln Phe Ala Lys Arg Leu Leu Lys Ala Gly Thr
170 20 25 30
172 Asp Val Thr Phe Phe Thr Ser Val Tyr Ala Trp Arg Arg Met Ala Asn
173 35 40 45
175 Thr Ala Ser Ala Ala Ala Gly Asn Pro Pro Gly Leu Asp Phe Val Ala
176 50 55 60
178 Phe Ser Asp Gly Tyr Asp Asp Gly Leu Lys Pro Cys Gly Asp Gly Lys
179 65 70 75 80
181 Arg Tyr Met Ser Glu Met Lys Ala Arg Gly Ser Glu Ala Leu Arg Asn
182 85 90 95
184 Leu Leu Leu Asn Asn His Asp Val Thr Phe Val Val Tyr Ser His Leu
185 100 105 110
187 Phe Ala Trp Ala Ala Glu Val Ala Arg Glu Ser Gln Val Pro Ser Ala
188 115 120 125
190 Leu Leu Trp Val Glu Pro Ala Thr Val Leu Cys Ile Tyr Tyr Phe Tyr
191 130 135 140
193 Phe Asn Gly Tyr Ala Asp Glu Ile Asp Ala Gly Ser Asp Glu Ile Gln
194 145 150 155 160
196 Leu Pro Arg Leu Pro Pro Leu Glu Gln Arg Ser Leu Pro Thr Phe Leu
197 165 170 175
199 Leu Pro Glu Thr Pro Glu Arg Phe Arg Leu Met Met Lys Glu Lys Leu
200 180 185 190
202 Glu Thr Leu Asp Gly Glu Glu Lys Ala Lys Val Leu Val Asn Thr Phe
203 195 200 205

```

RAW SEQUENCE LISTING                      DATE: 05/24/2000  
 PATENT APPLICATION: US/09/147,955        TIME: 10:58:35

Input Set : A:\1560-350.app  
 Output Set: N:\CRF3\05242000\I147955.raw

```

205 Asp Ala Leu Glu Pro Asp Ala Leu Thr Ala Ile Asp Arg Tyr Glu Leu
206      210      215      220
208 Ile Gly Ile Gly Pro Leu Ile Pro Ser Ala Phe Leu Asp Gly Gly Asp
209 225      230      235      240
211 Pro Ser Glu Thr Ser Tyr Gly Gly Asp Leu Phe Glu Lys Ser Glu Glu
212      245      250      255
214 Asn Asn Cys Val Glu Trp Leu Asp Thr Lys Pro Lys Ser Ser Val Val
215      260      265      270
217 Tyr Val Ser Phe Gly Ser Val Leu Arg Phe Pro Lys Ala Gln Met Glu
218      275      280      285
220 Glu Ile Gly Lys Gly Leu Leu Ala Cys Gly Arg Pro Phe Leu Trp Met
221      290      295      300
223 Ile Arg Glu Gln Lys Asn Asp Asp Gly Glu Glu Glu Glu Glu Leu
224 305      310      315      320
226 Ser Cys Ile Gly Glu Leu Lys Lys Met Gly Lys Ile Val Ser Trp Cys
227      325      330      335
229 Ser Gln Leu Glu Val Leu Ala His Pro Ala Leu Gly Cys Phe Val Thr
230      340      345      350
232 His Cys Gly Trp Asn Ser Ala Val Glu Ser Leu Ser Cys Gly Val Pro
233      355      360      365
235 Val Val Ala Val Pro Gln Trp Phe Asp Gln Thr Thr Asn Ala Lys Leu
236      370      375      380
238 Ile Glu Asp Ala Trp Gly Thr Gly Val Arg Val Arg Met Asn Glu Gly
239 385      390      395      400
241 Gly Gly Val Asp Gly Ser Glu Ile Glu Arg Cys Val Glu Met Val Met
242      405      410      415
244 Asp Gly Gly Glu Lys Ser Lys Leu Val Arg Glu Asn Ala Ile Lys Trp
245      420      425      430
247 Lys Thr Leu Ala Arg Glu Ala Met Gly Glu Asp Gly Ser Ser Leu Lys
248      435      440      445
250 Asn Leu Asn Ala Phe Leu His Gln Val Ala Arg Ala
251      450      455      460
254 <210> SEQ ID NO: 3
255 <211> LENGTH: 1474
256 <212> TYPE: DNA
257 <213> ORGANISM: Perilla frutescens
259 <220> FEATURE:
260 <221> NAME/KEY: CDS
261 <222> LOCATION: (29)..(1357)
263 <400> SEQUENCE: 3
264 accaaaccaa aacaaaattt ccacaaaa atg gtc cgc cgc cgc gtg ctg cta      52
265      Met Val Arg Arg Val Leu Leu
266      1      5
268 gca acg ttt ccg gcg caa ggc cac ata aat ccc gcc ctg caa ttc gcc      100
269 Ala Thr Phe Pro Ala Gln Gly His Ile Asn Pro Ala Leu Gln Phe Ala
270      10      15      20
272 aag aga ctg cta aaa gcc ggc act gac gtc acg ttt ttc acg agc gtt      148
273 Lys Arg Leu Leu Lys Ala Gly Thr Asp Val Thr Phe Phe Thr Ser Val
274 25      30      35      40

```

RAW SEQUENCE LISTING  
 PATENT APPLICATION: US/09/147,955  
 DATE: 05/24/2000  
 TIME: 10:58:35

Input Set : A:\1560-350.app  
 Output Set: N:\CRF3\05242000\I147955.raw

```

276 tat gca tgg cgc cgc atg gcc aac aca gcc tcc gcc gct gcc gga aac 196
277 Tyr Ala Trp Arg Arg Met Ala Asn Thr Ala Ser Ala Ala Ala Gly Asn
278                               45                               50                               55
280 cca ccg ggc ctc gac ttc gtg gcg ttc tcc gac ggc tac gac gac ggg 244
281 Pro Pro Gly Leu Asp Phe Val Ala Phe Ser Asp Gly Tyr Asp Asp Gly
282                               60                               65                               70
284 ctg aag ccc ggc ggc gac ggg aag cgc tac atg tcc gag atg aaa gcc 292
285 Leu Lys Pro Gly Gly Asp Gly Lys Arg Tyr Met Ser Glu Met Lys Ala
286                               75                               80                               85
288 cgc ggc tcc gag gcc tta aga aac ctc ctt ctc aac aac gac gac gtc 340
289 Arg Gly Ser Glu Ala Leu Arg Asn Leu Leu Leu Asn Asn Asp Asp Val
290                               90                               95                               100
292 act ttc gtc gtc tac tcc cac ctc ttt gca tgg gcg gag gtg gcg 388
293 Thr Phe Val Val Tyr Ser His Leu Phe Ala Trp Ala Ala Glu Val Ala
294 105                               110                               115                               120
296 cgt ttg tcc cac gtc ccg acc gcc ctt ctc tgg gtc gag ccc gcc acc 436
297 Arg Leu Ser His Val Pro Thr Ala Leu Leu Trp Val Glu Pro Ala Thr
298                               125                               130                               135
300 gtg ctg tgc ata tac cac ttc tac ttc aac ggc tac gca gac gag atc 484
301 Val Leu Cys Ile Tyr His Phe Tyr Phe Asn Gly Tyr Ala Asp Glu Ile
302                               140                               145                               150
304 gac gcc ggt tcc aat gaa att cag ctc cct cgc ctt cca tcc ctg gag 532
305 Asp Ala Gly Ser Asn Glu Ile Gln Leu Pro Arg Leu Pro Ser Leu Glu
306                               155                               160                               165
308 cag cgc agt ctt ccg acg ttt ctg ctg cct gcg acg ccg gag aga ttc 580
309 Gln Arg Ser Leu Pro Thr Phe Leu Leu Pro Ala Thr Pro Glu Arg Phe
310                               170                               175                               180
312 cgg ttg atg atg aag gag aag ctg gaa act tta gac ggt gaa gag aag 628
313 Arg Leu Met Met Lys Glu Lys Leu Glu Thr Leu Asp Gly Glu Glu Lys
314 185                               190                               195                               200
316 gcg aaa gta ttg gtg aac acg ttt gat gcg ttg gag ccc gat gca ctc 676
317 Ala Lys Val Leu Val Asn Thr Phe Asp Ala Leu Glu Pro Asp Ala Leu
318                               205                               210                               215
320 acg gct att gat agg tat gag ttg atc ggg atc ggg ccg ttg att ccc 724
321 Thr Ala Ile Asp Arg Tyr Glu Leu Ile Gly Ile Gly Pro Leu Ile Pro
322                               220                               225                               230
324 tcc gcc ttc ttg gac ggc gaa gat ccc tcc gaa acg tct tac ggc ggc 772
325 Ser Ala Phe Leu Asp Gly Glu Asp Pro Ser Glu Thr Ser Tyr Gly Gly
326                               235                               240                               245
328 gat ctt ttc gaa aaa tcg gag gag aat aac tgc gtg gag tgg ttg aac 820
329 Asp Leu Phe Glu Lys Ser Glu Glu Asn Asn Cys Val Glu Trp Leu Asn
330                               250                               255                               260
332 tcg aag ccg aaa tct tcg gtg gtg tat gtg tcg ttt ggg agc gtt ttg 868
333 Ser Lys Pro Lys Ser Ser Val Val Tyr Val Ser Phe Gly Ser Val Leu
334 265                               270                               275                               280
336 agg ttt cca aag gca caa atg gaa gag att ggg aaa ggg cta tta gcc 916
337 Arg Phe Pro Lys Ala Gln Met Glu Glu Ile Gly Lys Gly Leu Leu Ala
338                               285                               290                               295
340 tgc gga agg ccc ttt tta tgg atg ata cga gaa cag aag aat gac gac 964

```

RECEIVED

JUN -9 2000

TC 1600 MAIL ROOM

**Please Note:**

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

## VERIFICATION SUMMARY

DATE: 05/24/2000

PATENT APPLICATION: US/09/147,955

TIME: 10:58:36

Input Set : A:\1560-350.app

Output Set: N:\CRF3\05242000\I147955.raw

L:759 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:7  
L:888 M:258 W: Mandatory Feature missing, <220> not found for SEQ ID#:8  
L:888 M:258 W: Mandatory Feature missing, <221> not found for SEQ ID#:8  
L:888 M:258 W: Mandatory Feature missing, <222> not found for SEQ ID#:8  
L:888 M:258 W: Mandatory Feature missing, <223> not found for SEQ ID#:8  
L:888 M:340 W: (46) "n" or "Xaa" used: Feature required, for SEQ ID#:8  
L:891 M:258 W: Mandatory Feature missing, <220> not found for SEQ ID#:8  
L:891 M:258 W: Mandatory Feature missing, <221> not found for SEQ ID#:8  
L:891 M:258 W: Mandatory Feature missing, <222> not found for SEQ ID#:8  
L:891 M:258 W: Mandatory Feature missing, <223> not found for SEQ ID#:8  
M:340 Repeated in SeqNo=8